Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

10 November 2017

English only

Geneva, 13–17 November 2017
Item 6 of the revised provisional agenda
Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention

Autonomy in Weapon Systems

Submitted by the United States of America

I. Introduction

1. The United States of America acknowledges both the challenges and opportunities presented by emerging technologies in the area of lethal autonomous weapons systems (LAWS). As an overall matter, we believe that the law of war (also called international humanitarian law) provides a robust and appropriate framework for the regulation of all weapons in relation to armed conflict.

2. This working paper seeks to contribute to the meetings of the Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS) in Geneva between November 13-17, 2017, by providing the views of the United States on: (i) the legal review of weapons with autonomous functions in acquisition or development; (ii) the potential of weapon systems with autonomous functions to improve the implementation of law of war principles in military operations; and (iii) legal accountability regarding weapons with autonomous functions.

II. Legal review of weapons with autonomous functions in acquisition or development

3. The United States of America views the review of the legality of weapons as a best practice for implementing customary and treaty law relating to weapons and their use in armed conflict. The United States of America is not a party to the 1977 Additional Protocol I to the 1949 Geneva Conventions and therefore is not bound by that instrument, but we note that Article 36 of that Protocol creates an obligation for its Parties with respect to the study, development, acquisition, or adoption of a "new" weapon, means, or method of warfare.

4. Under United States Department of Defense (DoD) policies, legal reviews are conducted as part of the broader acquisition processes. In particular, a DoD policy (DoD Directive 5000.01) requires that the acquisition and procurement of DoD weapons and weapon systems be consistent with all applicable domestic and international law, including the law of war. To implement this requirement, DoD Directive 5000.01 provides that "[a]n

attorney authorized to conduct such legal reviews in the Department shall conduct the legal review of the intended acquisition of weapons or weapons systems."

5. Legal advice on the intended acquisition of weapons ordinarily may be provided at different stages in the process of acquiring such weapons. For example, under DoD policy on the use of autonomy in weapon systems (DoD Directive 3000.09), legal review and advice are provided before formal development and again before fielding with regard to certain types of weapon systems that involve the use of autonomy.

6. For the United States Department of Defense, the legal review of the acquisition or procurement of a weapon generally focuses on whether the weapon is illegal *per se* — whether a treaty to which the United States is a party or customary international law has prohibited its use in all circumstances. The use of autonomy to aid in the operation of weapons is not illegal *per se*.

7. Law of war issues related to targeting generally are not determinative of the lawfulness of a weapon. A legal review of a weapon should consider whether the weapon is "inherently indiscriminate," *i.e.*, whether the weapon is capable, under any set of circumstances and in particular the intended concept of employment, of being used in accordance with the principles of distinction and proportionality. Nevertheless, most targeting issues (*e.g.*, whether a weapon would be used consistent with the requirement that attacks may only be directed against military objectives) are only capable of determination when presented with the facts of a particular military operation.

8. Weapons that use autonomy in target selection and engagement seem unique in the degree to which they would allow consideration of targeting issues during the weapon's development. For example, if it is possible to program how a weapon will function in a potential combat situation, it may be appropriate to consider the law of war implications of that programming. In particular, it may be appropriate for weapon designers and engineers to consider measures to reduce the likelihood that use of the weapon will cause civilian casualties.

9. Under DoD policy, autonomous and semi-autonomous weapons systems go through "rigorous hardware and software verification and validation (V&V) and realistic system developmental and operational test and evaluation (T&E)." Although rigorous testing and sound development of weapons are not required by the law of war as such, these good practices can support the implementation of law of war requirements. Rigorous and realistic testing standards and procedures can ensure that commanders and national security policy makers can have a reasonable expectation of the likely effects of employing the weapon in different operational contexts. In addition, such practices can help reduce the risk of unintended combat engagements, such as weapons malfunctions that could inadvertently cause harm to civilians.

III. Compliance with the law of war in using weapon systems with autonomous functions

10. The law of war rules on conducting attacks (*e.g.*, the rules relating to distinction and proportionality) impose obligations on States and other parties to a conflict, and it is for individual human beings, commensurate with their role within the State or party to the conflict, to ensure compliance with those obligations when employing any weapon or weapons system, including autonomous or semi-autonomous weapons systems. For example, DoD policy recognizes that "[p]ersons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must," among other requirements, "do so … in accordance with the law of war."

11. It is not the case that the law of war requires that a weapon, even a semi-autonomous or autonomous weapon, make legal determinations. For example, the law of war does not require that a weapon determine whether its target is a military objective, but rather that the weapon be capable of being employed consistent with the principle of distinction. Similarly, the law of war does not require that a weapon make proportionality determinations, such as whether an attack is expected to result in incidental harm to

civilians or civilian objects that is excessive in relation to the concrete and direct military advantage expected to be gained.

12. The law of war does not require weapons to make legal determinations, even if the weapon (*e.g.*, through computers, software, and sensors) may be characterized as capable of taking some form of action or decision in a given moment in the absence of direction by a human being, such as whether to fire the weapon or to select and engage a target. Relatively rudimentary autonomous weapons, such as homing missiles, have been employed for many years, and there has never been a requirement that such weapons themselves determine that legal requirements are met.

13. Rather, it is persons who must comply with the law of war by employing weapons in a discriminate and proportionate manner. For example, even if the weapon autonomously selects and engages targets, its use would be precluded when the use of the weapon would be expected to result in incidental harm to civilians or civilian objects that is excessive in relation to the concrete and direct military advantage expected to be gained.

14. In addition, the obligation to take feasible precautions in order to reduce the risk of harm to civilians and other persons or objects protected from being made the object of attack must be considered when using weapon systems with advanced autonomous functions. For example, depending on the circumstances, it might be feasible to monitor the operation of the weapon system and to stop its operation in the event that it malfunctioned or the circumstances change. As another example, it might be appropriate to consider whether it is possible to program or build mechanisms into the weapon that would reduce the risk of civilian casualties while in no way decreasing the military advantages offered by the weapon. A best practice in this regard may be found in the requirements in DoD policy for the interface between people and machines for autonomous and semi-autonomous weapons to: (1) be readily understandable to trained operators; (2) provide traceable feedback on system status; and (3) provide clear procedures for trained operators to activate and deactivate system functions. These requirements to improve human-machine interfaces assist operators in making accurate judgments regarding the use of force.

15. The ability of weapons to make decisions or assessments of issues that would be considered under law of war can be viewed as an additional feature that improves the ability of human beings to implement legal requirements rather than as an effort to replace a human being's responsibility and judgment under the law.

IV. Potential for autonomy in weapon systems to improve the implementation of law of war principles in military operations

16. In many cases, the use of autonomy in weapon systems could enhance the way law of war principles are implemented in military operations.

17. For example, very basic applications of autonomy allow some munitions to self-deactivate or to self-destruct, which helps reduce the risk these weapons may pose to the civilian population after the munitions have served their military purpose.

18. More advanced applications of autonomy may facilitate greater precision in guidance of bombs and missiles against military objectives, reducing the likelihood of inadvertently striking civilians and civilian objects as compared to the use of unguided bombs and missiles to achieve the same desired result.

19. Similarly, autonomous functions allow defensive systems to select and engage incoming enemy projectiles, such as mortars, artillery shells, and rockets. These defensive systems can provide military commanders more time to decide on how to respond to the threat. For example, directing "counter-battery fire" against the origin of the enemy projectiles has been a common response to such attacks, and the additional time afforded by autonomous defensive systems could allow military commanders more time to consider and execute a more deliberate and precise response.

20. These applications of autonomy illustrate a fundamental feature of the law of war — the law of war often reflects the convergence of military and humanitarian interests.

21. Autonomy can be used in weapon systems to create more capabilities. Commanders can use additional capabilities to increase the efficiency of military operations — more precisely applying force and causing less unintended destruction. Improving efficiency is done for sound military reasons — to allow fewer resources to accomplish more military purposes. But the same capabilities that reduce wasteful or incorrect applications of military force, such as incidents of "friendly fire," can also reduce the risk of civilian casualties.

22. For example, militaries might develop weapons with advanced technologies, such as smart grenade launchers, to give their soldiers new advantages in countering the use of cover by enemy fighters to avoid small arms fire. By reducing the need for even greater applications of force such as artillery or air bombardments, these weapons have potentially long-term benefits by reducing the effects of larger explosive weapons in populated areas or the presence of explosive remnants of war.

23. These types of "smart" weapons might create additional options for commanders — allowing attacks to be conducted in circumstances where the use of "dumb" weapons would cause significant or excessive civilian casualties. This, however, should not be construed as necessarily requiring States to use "smart" weapons when available rather than "dumb" weapons.

24. It is expected that further developments in autonomous and semi-autonomous weapon systems will allow military forces to apply force more precisely and with less collateral damage than would be possible with existing systems.

V. Legal accountability and weapons with autonomous functions

25. Machines are not States or persons under the law. Questions of legal accountability are questions of how existing and well-established principles of State and individual responsibility apply to States and persons who use weapon systems with autonomous functions.

26. As a general principle, States are responsible for the acts of persons forming part of their armed forces. It follows that States are responsible for the uses of weapons with autonomous functions by persons forming part of their armed forces as well as other such acts that may be attributable to a State under the law of State responsibility. States, in ensuring accountability for such conduct, may use a variety of mechanisms, including investigations, individual criminal liability, civil liability, and international disciplinary measures.

27. As with all decisions to employ weapon systems, persons are responsible for their individual decisions to use weapons with autonomous functions. For example, persons who use weapons with autonomous functions to violate the prohibition on targeting the civilian population may be held responsible for such violations.

28. The responsibilities of any particular individual belonging to a State or a party to the conflict may depend on that person's role in the organization or military operations. As a general matter, the persons who are responsible for implementing a party to a conflict's obligation are those persons with the authority to make the necessary decisions and judgments required by that international obligation. For example, a party to a conflict has the obligation to take feasible precautions to reduce the risk to civilians, such as providing warnings before attacks. The determination of whether it is feasible to provide such a warning would be made by the relevant commander in charge of the attack.

29. As noted above, advanced applications of autonomy in weapon systems can allow for issues that would normally only be presented in the context of the use of the weapon system to be presented in the context of the development of the weapon system. Persons who engage in wrongdoing in the development and testing of a weapon could be held accountable, at least under principles and rules of accountability in domestic law.

30. Intentional wrongdoing involving weapons is clearly prohibited. In the absence of intentional wrongdoing, assessments of accountability may be more complex. Mere accidents or equipment malfunctions are not violations of the law of war, even if civilians are killed or injured as a result of those malfunctions. The standard of care or regard that is due in conducting military operations with regard to the protection of civilians is a complex question to which the law of war does not provide a simple answer. This standard must be assessed based on the general practice of States and common standards of the military profession in conducting operations.

31. A general principle of accountability, which is reflected in the law of war, is that decision-makers must be judged based on the information available to them at the time and not on the basis of information that subsequently comes to light. Thus, for example, in assessing whether a commander's decision to use weapons with autonomous functions was reasonable in a particular context, whether the commander acted in good faith based on the information available to him or her at the time would need to be considered. In this regard, training on the weapon system and rigorous testing of the weapon system can help commanders be advised of the likely effects of employing the weapon system. These measures, found in DoD policy, can help promote good decision-making and accountability.

Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

10 November 2017

English only

Geneva, 13–17 November 2017
Item 6 of the revised provisional agenda
Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention

Characteristics of Lethal Autonomous Weapons Systems

Submitted by the United States of America

I. Introduction

1. This working paper seeks to contribute to the meetings of the Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS) in Geneva between November 13–17, 2017, by providing the views of the United States on: (i) identifying characteristics of "LAWS" rather than negotiating a "definition" in these GGE discussion; (ii) recommendations regarding characteristics of LAWS; and (iii) definitions used by the United States Department of Defense in internal policies on the use of autonomy in weapon systems.

II. Identifying characteristics of LAWS rather than negotiating a definition of "LAWS" in these GGE discussions

2. The United States believes that it is unnecessary for the GGE to adopt a specific working definition of LAWS. Instead, we support promoting a general understanding of the characteristics of LAWS. We believe that the absence of a specific working definition is no impediment to the GGE's work in understanding the potential issues posed by LAWS. Given that the law of war provides a robust and coherent system of regulation for the use of weapons, the GGE can discuss the issues potentially posed by "LAWS" under the object and purpose of the CCW without needing to agree on a specific working definition of LAWS. For example, as explained in the United States working paper on legal issues, the law of war's existing rules of general applicability apply with respect to the use of all weapons, including any weapons deemed to be "LAWS."

3. A legal definition is generally developed for the specific purposes of a legal rule and not in the abstract. Often legal definitions determine the scope of a legal rule, *i.e.*, the matters to which the rule would apply. For example, the definition of "remotely-delivered mine" in the CCW Amended Protocol II identifies what types of mines are subject to a set of restrictions in that Protocol.

4. A working definition should not be drafted with a view toward describing weapons that should be banned. This would be premature and counterproductive because it would divert time and effort from understanding the issues to negotiating what would be covered.

GE.17-19939(E)
1719939

Please recycle

As the High Contracting Parties have not decided to negotiate or adopt a new protocol specifically to ban or regulate LAWS, any common understanding of LAWS must not prejudice future decisions regarding potential outcomes.

5. In identifying characteristics of LAWS, we must be cautious not to make hasty judgments about the value or likely effects of emerging or future technologies. Frequently, we may change our views of technologies over time as we gain more experience with them. In particular, we want to encourage innovation and progress in addressing the objects and purposes of the Convention.

III. Recommendations regarding characteristics of LAWS

6. Although we believe it unnecessary for the GGE to seek to negotiate a single working definition, we support identifying general characteristics of LAWS in order to promote our understanding of the relevant concepts or issues in these GGE discussions. Identifying general characteristics of LAWS will help us understand what is generally referred to by this term, without providing a definition that would establish the parameters or what is, or is not, included. This flexibility in approach is important given that scientists and engineers continue to develop new technological advancements and that our understanding continues to improve. In light of that purpose, we offer the following recommendations regarding characteristics of LAWS.

7. The characteristics of LAWS should be intelligible to all relevant audiences, including roboticists, engineers, scientists, lawyers, military personnel, and ethicists. The characteristics of LAWS should not be identified based on specific technological assumptions such that the characteristic would be rendered obsolete by technological developments. In this regard, we should not articulate specific levels of autonomy or types of machine reasoning. Our sense is that creating technical categories like this or seeking to define "artificial intelligence" would be especially ill-advised because there are already diverse taxonomies along these lines and because scientists and engineers continue to develop technological advancements.

8. Seeking to define the sophistication of the machine intelligence would incorrectly focus on the machine, rather than understanding what is important for the law — how human beings are using the weapon and what they expect it to do. For example, it is irrelevant under the law of war whether a rocket engine is powered by a solid fuel or a liquid propellant. Rather, the law of war is concerned with how that power is used in combat. Similarly, focusing on the sophistication of the "analytical engine" powering a weapon (e.g., what type of algorithm or method of machine learning is employed) risks ignoring the focus of the law — how humans will use that weapon (e.g., using the machine to select and engage targets without further intervention by a human operator).

9. Lastly, focusing on the machine also could stimulate unwarranted fears that are more the product of science fiction and popular imagination than fact.

IV. Definitions used by the United States Department of Defense in internal policies on the use of autonomy in weapon systems

10. In light of the above considerations and to further the GGE's understanding of some of the relevant concepts and issues related to the characteristics of LAWS, we offer for consideration definitions that the Department of Defense has specifically developed for use in its internal policies relating to the use of autonomy in weapon systems.

11. Although the GGE's purpose may be different than the purposes for which these definitions were created, we believe these definitions help demonstrate the careful thought that should be applied when identifying the relevant characteristics of LAWS. For example, these definitions were developed after considering existing weapon systems. These definitions do not depend on a technical characterization of the sophistication of the machine reasoning. Instead, these definitions focus on what we believe to be the most

important issues posed by the use of autonomy in weapon systems — people who employ these weapons can rely on the weapon systems to select and engage targets.

12. Department of Defense policy includes the following definitions:

(a) "**autonomous weapon system**. A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation."

(b) "**semi-autonomous weapon system.** A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator. This includes:

> Semi-autonomous weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement.
>
> 'Fire and forget' or lock-on-after-launch homing munitions that rely on TTPs [tactics, techniques, and procedures] to maximize the probability that the only targets within the seeker's acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator."

13. The United States Department of Defense Directive establishing these definitions "[d]oes not apply to autonomous or semi-autonomous cyberspace systems for cyberspace operations; unarmed, unmanned platforms; unguided munitions; munitions manually guided by the operator (e.g., laser- or wire-guided munitions); mines; or unexploded explosive ordnance." Therefore, these types of systems, platforms, weapons, or devices, would not be considered to fall within the rules established by this Department of Defense Directive for "autonomous" or "semi-autonomous" weapons.

Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

10 November 2017

English only

Geneva, 13–17 November 2017
Item 6 of the revised provisional agenda
Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention

Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention

Submitted by the Russian Federation

1. The Russian Federation participated in all three informal meetings of experts on lethal autonomous weapons systems (LAWS) held in May 2014, April 2015 and April 2016 within the framework of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (the Inhumane Weapons Convention, the CCW). In April 2016, the Russian Federation refrained from supporting the recommendation to the Fifth CCW Review Conference adopted at the third informal Meeting of Experts to hand this work over to an open-ended Group of Governmental Experts (GGE) on LAWS with a mandate for discussion, having set out its position as a dissenting opinion that it would be useful to continue with the informal format of discussions. During the actual CCW Review Conference (Geneva, 12–16 December 2016), however, the Russian Federation, guided by the basic principle not to block multilateral arms control discussions not directly undermining national defense and security, chose to break no consensus and refrained from opposing the establishment of the open-ended GGE on LAWS with a mandate for discussion within the framework of the Convention while adopting the Final Document of the CCW Review Conference.

2. The Russian Federation would like to clarify its position on LAWS, as well as to share some thoughts as to the upcoming formal work in this respect within the CCW.

I. Overall assessment of the work on LAWS within the CCW

3. The Russian Federation still feels positive of the decision to hand the discussions on LAWS over to the CCW platform. This Convention, being an important tool of international law that most fully ensures the balance between the humanitarian concerns and legitimate security interests of States, is the best format to deal with these issues.

4. In the opinion of the Russian Federation, the previous informal discussions on LAWS on the CCW platform were generally useful. This is particularly true for the two recent meetings of experts chaired by Germany where the discussions became more

structured. In particular, their most important deliverables include recognition by the States that the existing provisions of international humanitarian law are applicable to LAWS and human control over them should be continued.

5. The informal discussions, however, have essentially failed to advance the understanding of LAWS as a subject matter of possible accords and remove the remaining doubts as to the feasibility of an in-depth study of this topic. It is still ill-considered and contradictory. This, particularly, is proved by the fact that the discussions remain purely academic and not always easy for perception.

II. Persisting challenges in discussion on LAWS

6. According to the Russian Federation, the lack of working samples of such weapons systems remains the main problem in the discussion on LAWS. Certainly, there are precedents of reaching international agreements that establish a preventive ban on prospective types of weapons. However, this can hardly be considered as an argument for taking preventive prohibitive or restrictive measures against LAWS being a by far more complex and wide class of weapons of which the current understanding of humankind is rather approximate.

7. The frequently cited CCW Protocol IV on Blinding Laser Weapons can hardly be applied to the work on LAWS, since this document, as is known, prohibited not the laser weapons per se, but a very specific case of their employment — to cause permanent blindness.

8. The difficulty of making a clear distinction between civilian and military developments of autonomous systems based on the same technologies is still an essential obstacle in the discussion on LAWS. It is hardly acceptable for the work on LAWS to restrict the freedom to enjoy the benefits of autonomous technologies being the future of humankind.

III. Future vision

9. As for the future work on LAWS within the Group of Governmental Experts, the Russian Federation considers appropriate to focus on the following issues.

IV. Definition

10. We are sympathetic with the Food-for-thought Paper proposed by the Indian GGE Chairperson, as well as with the fact that in our work we continue to analyze this issue in its multiple dimensions, thus returning, to a certain extent, to the working methods of informal meetings of experts. We do indeed need to look seriously into the matter of LAWS in order to understand how to move forward.

11. At the same time, we cannot avoid considering the key issues without which further discussion on LAWS will face increasing practical difficulties. This refers to harmonizing the basic definitions of LAWS. Unfortunately, there are examples in the negotiation practice when key definitions were not the first to be developed. In particular, one of such examples is the Convention on Cluster Munitions (Oslo Process, CCM) prohibiting the use, stockpiling, production and transfer of cluster munitions. It should be recalled that the definition contained in this document took a selective approach to cluster munitions subject to destruction. As a result, these weapons were divided into "bad" and "good" on the basis of lame criteria. This subsequently became the factor keeping a number of States from joining the CCM.

V. Other concepts

12. Further work on key aspects of such weapons which relate to the notions of "autonomy", "critical functions", "meaningful human control," "predictability," etc., will depend on the definition of LAWS. The Russian Federation has no objection to parallel efforts to develop definitions with discussing key aspects of LAWS, as this would facilitate the issue of definition of such systems.

VI. Concept of "meaningful human control"

13. This concept, though poorly developed, is a potential element of consent. The overwhelming majority of States agree on the inadmissibility of loss of meaningful human control of such weapons systems. However, it should be recognized that it will be very difficult to develop criteria for the "meaningfulness" of such control without politicizing this issue.

VII. International legal aspects of LAWS

14. The Russian Federation believes that the GGE could conduct a thorough review of existing provisions of international law, including international humanitarian law and human rights law that could potentially be applied to LAWS, and express its opinion as to the adequacy of such provisions to address the existing concerns.

15. In general, the Russian Federation proceeds from the premise that the examination of new issues within the CCW should be carried out in the most careful, balanced and considered manner, taking into account both humanitarian concerns and legitimate defense interests of States. That said, the need to address humanitarian concerns cannot be used as the one and only sufficient prerequisite for imposing restrictive and prohibitive regimes on certain weapons.

Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

10 November 2017

English only

Geneva, 13–17 November 2017
Item 6 of the revised provisional agenda
Examination of various dimensions of emerging technologies in the area of lethal autonomous weapons systems, in the context of the objectives and purposes of the Convention

A "compliance-based" approach to Autonomous Weapon Systems

Working Paper submitted by Switzerland[1]

I. Introduction

1. Ever greater investments go into research and development in the area of robotics and artificial intelligence. This trend is likely to bring increasingly autonomous capabilities to weapon systems, which in turn bears the potential of redefining the role of humans in the use of force.

2. Switzerland is convinced that compliance with international law, notably international humanitarian law (IHL), must be central to discussions about autonomous weapon systems and should figure prominently in the report of the GGE as well as in the way forward.

3. The present working paper recapitulates requirements for compliance with IHL, reiterates considerations by Switzerland on a possible working definition of autonomous weapon systems, and on that basis, identifies elements for a "compliance-based" approach aimed at advancing the debate within the CCW in an inclusive and constructive manner.

II. Ensuring compliance with international humanitarian law

4. IHL is the most relevant body of international law governing the development of autonomous weapon systems and their employment in armed conflicts and will therefore be the main focus of this section. This being said, other branches of international law, most notably human rights law, may equally impose limits on the use of force in armed conflicts, and international criminal law governs individual criminal responsibility for violations. Moreover, any recourse to the threat or use of force by States is governed by the UN Charter.

[1] Adapted version of the working paper submitted to the Informal meeting of experts on lethal autonomous weapons systems (LAWS), Geneva, 11-15 April 2016. Changes were made with regard to the working definition in paragraphs 26 – 29 and the process in paragraph 37.

5. There is general consensus that the potential development and employment of any autonomous weapon systems must remain in compliance with existing international law and, in times of armed conflict, particularly with IHL. The International Court of Justice was clear in its 1996 Advisory Opinion that the established principles and rules of humanitarian law applicable in armed conflict apply to "all forms of warfare and to all kinds of weapons, those of the past, those of the present and those of the future".[2] Even though the treaty provisions regulating the conduct of hostilities do not expressly refer to new technologies of warfare as such, the development and employment of such technologies in armed conflict always must be in accordance with IHL. Customary IHL rules, in particular those related to the conduct of hostilities, equally apply to all means and methods of warfare. Indeed, it is a long standing principle that the right of parties to an armed conflict to choose methods or means of warfare is not unlimited.

6. Against this background, three distinct issues should be considered: (1) the substantive provisions of IHL applicable to the employment of autonomous weapon systems in armed conflict; (2) the corresponding procedural rule of IHL requiring the conduct of legal weapons reviews as a means of ensuring conformity with international law; and (3) the issue of accountability.

A. Substantive IHL provisions

7. At the outset, it is important to distinguish between the lawfulness of a particular type of weapon as such (weapons law), and the lawfulness of the way in which it is being used (targeting law). While every weapon can be used in an unlawful manner, the inherent characteristics of certain weapons categories entail that their use – in some or all circumstances – is unlawful *per se*. The permissibility of all other weapons depends on their employment in conformity with IHL.

8. Under IHL, any weapon possessing one or more of the following characteristics is inherently unlawful : (1) the weapon is of a nature to cause superfluous injury or unnecessary suffering; (2) the weapon is indiscriminate by nature because it cannot be aimed at a lawful target or because its effects cannot be limited as required by IHL; (3) the weapon is intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment; (4) the weapon has been specifically prohibited in treaty or customary law. These criteria apply to all weapons, including new technologies such as autonomous weapon systems.

9. Of particular relevance for autonomous weapon systems is the prohibition of indiscriminate weapons. A weapon system would have to be regarded as indiscriminate if it cannot be directed at a specific military objective or if its effects cannot be limited as required by IHL and if, in either case, it is of a nature to strike military objectives and civilians or civilian objects without distinction. In other words, in order for an autonomous weapon system to be lawful under this rule, it must be possible to ensure that its operation will not result in unlawful outcomes with respect to the principle of distinction.

10. With regard to the *lawful use* of a weapon system, the principles governing the conduct of hostilities need to be considered. Most notably, in order to lawfully use an autonomous weapon system for the purpose of attack, belligerents must: (1 - Distinction) distinguish between military objectives and civilians or civilian objects and, in case of doubt, presume civilian status; (2 - Proportionality) evaluate whether the incidental harm likely to be inflicted on the civilian population or civilian objects would be excessive in relation to the concrete and direct military advantage anticipated from that particular attack; (3 - Precaution) take all feasible precautions to avoid, and in any event minimize, incidental harm to civilians and damage to civilian objects; and cancel or suspend the attack if it becomes apparent that the target is not a military objective, or that the attack may be expected to result in excessive incidental harm.

[2] International Court of Justice, Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion of 8 July 1996, para. 86.

11. The employment of autonomous weapon systems in the conduct of hostilities also raises particular challenges with regard to the prohibition of the denial of quarter and the protection of persons *hors de combat*, i.e. the protection from attack of the wounded and sick and those intending to surrender.[3] Any reliance on autonomous weapon systems would need to preserve a reasonable possibility for adversaries to surrender. A general denial of this possibility would violate the prohibition of ordering that there shall be no survivors or of conducting hostilities on this basis (denial of quarter).[4]

12. Full compliance with IHL, however, is not limited to the rules governing the conduct of hostilities. Besides employing autonomous weapon systems as a weapon in attack, it is also conceivable that such systems could be used to perform other tasks governed by IHL, such as the guarding and transport of persons deprived of their liberty or tasks related to crowd control and public security in occupied territories. Additional specific rules need to be taken into consideration if autonomous weapon systems were to be relied on for such activities

13. Applying these requirements of *lawful use* to autonomous weapon systems is not without complexity. The Geneva Conventions of 1949 and their Additional Protocols of 1977 were undoubtedly conceived with States and individual humans as agents for the exercise and implementation of the resulting rights and obligations in mind. In addition, many pivotal rules of IHL presume the application of evaluative decisions and value judgements, such as the presumption of civilian status in case of "doubt",[5] the assessment of "excessiveness" of expected incidental harm in relation to anticipated military advantage, the betrayal of "confidence" in IHL in relation to the prohibition of perfidy, and the prohibition of destruction of civilian property except where "imperatively" demanded by the necessities of war.[6] The principle of precaution even expressly refers to "those who plan or decide upon" an attack,[7] and the provisions establishing criminal responsibility for serious violations of IHL also are based on a manifest presumption of human agency.[8]

14. Furthermore, the overarching obligation of all belligerents to "respect and ensure respect" for IHL "in all circumstances" seems to imply a derived duty of exercising sufficient control or supervision over the development and/or employment of autonomous weapon systems to ensure full compliance with IHL and prevent outcomes that would be unlawful under existing international law. In accordance with this obligation, it is uncontested that preparatory measures must be taken to permit the implementation of IHL and implementation should be supervised.[9]

15. Accordingly, given the current state of robotics and artificial intelligence, it is difficult today to conceive of an autonomous weapon system that would be capable of reliably operating in full compliance with *all* the obligations arising from existing IHL without any human control in the use of force, notably in the targeting cycle.

16. On this basis, the question, therefore, is not whether States have a duty to control or supervise the development and/or employment of autonomous weapon systems, but how that control or supervision ought to be usefully defined and exerted. Would it be sufficient, for example, to rely on superior programming and strict reliability testing to make an autonomous weapon system predictably compliant with IHL for its intended operational

---

[3] Article 41 of Additional Protocol I to the Geneva Conventions.

[4] Article 40 of Additional Protocol I to the Geneva Conventions.

[5] Article 50(3) and article 52(3) of Additional Protocol I to the Geneva Conventions.

[6] Article 23(g) of Hague Regulation IV, Article 53 of the Fourth Geneva Convention.

[7] Article 57(2)(a) of Additional Protocol I to the Geneva Conventions.

[8] Article 49/50/129/146 Geneva Conventions; Section III of Additional Protocol I to the Geneva Conventions.

[9] Cf. Yves Sandoz, Christophe Swinarski, Bruno Zimmermann (eds.), Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, ICRC, 1987, Article 1 Additional Protocol I, para 41. See further article 80 of Additional Protocol I to the Geneva Conventions, which requires that the High Contracting Parties and the parties to the conflict shall "take all necessary measures for the execution of their obligations", "give orders and instructions to ensure observance" and "supervise their execution".

parameters? If so, would it be permissible to restrict human involvement to the proper activation of such an autonomous weapon system? This working paper does not seek to prejudge these questions.

17. However, it is useful to recognize that control can be exercised in various different ways, both independently and in combination. Arguably, in the future, a significant level of control can already be exerted in the development and programming phase. Through testing and evaluating autonomous weapon systems in the course of weapons reviews, predictability and reliability of such systems can also be reinforced. Predictability and reliability can also be increased by restricting the autonomous weapon systems' parameters of engagement in line with the system's compliance capabilities. Depending on operational requirements and system capabilities, further control can be exercised through real-time supervision, or through an autonomous or human operated override mechanism aimed at avoiding malfunction or, alternatively, ensuring safe failure. Assessing, amongst others, these parameters in relation with existing weapon systems may contribute to a better understanding of appropriate means of ensuring compliance with IHL in the development and use of autonomous weapon systems for military purposes. As an illustration, those current weapon systems that are expected to operate with limited human control, such as certain booby-traps, are generally subject to severe restrictions as to the geographic area and mode of their use. The applicability of such restrictions with regard to systems with far greater and more sophisticated autonomy could be assessed in order to understand their relevance in terms of IHL compliance.

18. This overview of the relevant rules of IHL would be incomplete without reference to the Martens clause. The Martens Clause, which is part of customary international law, affords an important fallback protection in as much as the "laws of humanity and the requirements of the public conscience" need to be referred to if IHL is not sufficiently precise or rigorous. [10] Accordingly, not everything that is not explicitly prohibited can be said to be legal if it would run counter the principles put forward in the Martens clause. Indeed, the Martens clause may be said to imply positive obligations where contemplated military action would result in untenable humanitarian consequences.

B. Legal reviews

19. Under IHL, the substantive rules described above are complemented by a procedural rule. As with any other weapon, means or method of warfare, States have the positive obligation to determine, in the study, development, acquisition or adoption of any autonomous weapon system, whether their employment would, in some or all circumstances, contravene existing international law. In this regard the duty to conduct legal reviews, as specified in article 36 of Additional Protocol I to the Geneva Conventions, constitutes an important element in preventing or restricting the development and employment of new weapons that would not meet the obligations listed above.[11] Moreover, adequate testing and reviews may also have implications on the level of State responsibility, including for malfunction of approved autonomous weapon systems.

20. The legal review of autonomous weapon systems may present a number of challenges distinct from traditional weapons reviews. Specifically, the question is how such

[10] The "principles of humanity and the dictates of public conscience" are contained in the so called Martens Clause, which is repeated notably in the preamble of the CCW and in Art. 1(2) of Additional Protocol I. In its 1996 Advisory Opinion on the legality of the threat or use of nuclear weapons, the International Court of Justice held that the clause "proved to be an effective means of addressing the rapid evolution of military technology" (§78).

[11] ICRC, A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977 (2006), available at https://www.icrc.org/eng/assets/files/other/icrc_002_0902.pdf. See also the working paper submitted by the Netherlands and Switzerland to the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS), Weapons Review Mechanisms (CCW/GGE.1/2017/WP.5), available at: http://undocs.org/ccw/gge.1/2017/WP.5.

systems and their specific characteristics can be meaningfully tested. Beyond the purely technical challenge of assessing IHL compliance of an autonomous weapon system, there is also a conceptual challenge related to the fact that an autonomous system will assume an increasing number of determinations in the targeting cycle which traditionally are being taken care of by a human operator. For example, in traditional systems, the principle of proportionality was to be respected by the operator. It consequently fell outside the scope of an article 36 review. However, if an autonomous weapon system is expected to perform this proportionality assessment by itself, that aspect will need to be added to legal reviews of these systems. New evaluation and testing procedures may need to be conceptualized and developed to meet this particular challenge.

21. Given the special characteristics of autonomous weapon systems, a number of further measures could be recommended for incorporation into national review procedures. For example, one could imagine recommending that in some cases, particular safeguards against malfunction, such as the possibility of a human override, are built into autonomous weapon systems. Proper understanding of a system's predictability, especially when it comes to interaction with other autonomous systems, could also be named as example. While the process of national legal reviews may require procedural and technical adaptations to fully capture the complexity of autonomous weapon systems, if rigorously implemented, it holds the potential of ensuring that all new weapons, means and methods of warfare are developed and acquired in compliance with international law.

C. Accountability

22. Another important issue arising with regard to autonomous weapon systems is that of accountability, namely in terms of individual criminal responsibility and of state responsibility. Given that autonomous weapon systems possess no agency or legal personality of their own, the question of individual criminal responsibility focuses entirely on the responsibility of humans that are involved as operators, commanding officers, programmers, engineers, technicians or in other relevant functions. If the deployment of an autonomous weapon system results in a serious violation of IHL, and if that violation is the consequences of culpable fault on the part of a human being the latter may be subjected to criminal prosecution for war crimes or, depending on the circumstances of the case, also for crimes against humanity or genocide.

23. Criminal culpability is self-evident in the case of deliberate and premeditated intent. It is less so in the case of recklessness or (advertent) negligence, or of simple acceptance of a risk that violations will or may occur. With regard to war crimes, article 85(3) of Additional Protocol I requires "willfulness", with national practices varying as to the meaning to be given to this requirement. The International Tribunal for the Former Yugoslavia has stated that, as a matter of customary law, indirect intent would be sufficient to fulfil the mental requirement *(mens rea)*.[12] Conversely, the Rome Statute of the International Criminal Court does not foresee criminal liability for indirect intent,[13] except in the case of command responsibility for the conduct of subordinates.[14] As a matter of concept, command responsibility does not entail the commander's direct criminal responsibility for crimes committed by his subordinates, but for his or her culpable failure to prevent, suppress or repress crimes committed by persons (i.e. not machines) under his or her command and control. Strictly speaking, therefore, a commander's failure to duly control autonomous weapon systems operating under his command is not a case of command responsibility within the contemporary understanding of this concept, but may constitute a direct violation of the duties of precaution, distinction, proportionality or any other obligation imposed by IHL. This does not exclude that, as the functions of human soldiers are increasingly "delegated" to autonomous weapon systems, it may become appropriate *de lege ferenda* to extend the commander's supervisory duty, *mutatis mutandis*

[12] Cf. ICTY, Prosecutor v. Tihomir Blaskic, Judgement of 29 July 2004, Appeals Chamber, para. 42.
[13] Article 30 of the Rome Statute of the International Criminal Court.
[14] Article 28 of the Rome Statute of the International Criminal Court.

and *by analogy*, also to autonomous weapon systems operating under his direct command and control.

24. Overall, under current international law, whether or not there is an "accountability gap" for operators, commanders and other humans involved in the operation of autonomous weapon systems depends on the applicable *mens rea* standard. As a general assumption, the more significant human involvement in a specific autonomous weapon system operation is (such as humans "in the loop"), the easier it is to assign individual responsibility. This assumption may be relevant with a view to the general obligation of States to respect and ensure respect for IHL.

25. The second dimension of accountability derives from general international law governing the responsibility of States for internationally wrongful acts. States remain legally responsible for unlawful acts and resulting harm caused by autonomous weapon systems they employ, including due to malfunction or other undesired or unexpected outcomes. The rules governing *attribution* of conduct to a State are pertinent in relation to autonomous weapon systems as with any other means and methods of warfare. Given that autonomous weapon systems lack legal personality in the first place, they cannot become agents in a human sense, whether state agents or non-state actors. The question of State responsibility therefore does not turn on the nature or capability of the autonomous weapon systems, but of legal and factual status of the person or entity deciding on its employment. A decision of a person or entity exercising public powers or governmental authority (e.g. the armed forces) to employ an autonomous weapon system in a given situation certainly would be attributable to the State.[15] The result is that States cannot escape international responsibility by a process of "delegating" certain tasks to autonomous weapon systems.

III. Working definition

26. Switzerland remains of the view that, at the current stage, it is premature to aim for a *definition* of (lethal) autonomous weapon systems that seeks to draw a line between desirable, acceptable or unacceptable systems. However, Switzerland considers it relevant that the GGE endeavours to develop a common understanding of both the substance and the purpose of a *preliminary working definition*.

27. With regard to substance, such a working definition should provide for a shared understanding that there is a need to focus on systems with elements of autonomy in the targeting cycle. In addition, this working paper considers that the element of lethality, though of particular concern in practice, should not be conceptually regarded as a prerequisite characteristic of autonomous weapon systems. It prefers a more inclusive understanding of autonomous weapon systems, which would also cover means and methods of warfare that do not necessarily inflict physical death, but the effects of which may be restricted to causing, for example: (1) physical injury short of death, (2) physical destruction of objects, or (3) non-kinetic effects.

28. With regard to the purpose of the working definition, Switzerland is of the view that such a working definition should not prejudge the question of which systems may potentially require a regulatory response. Keeping in mind that fully autonomous weapon systems do not yet exist, it should allow for the examination of all relevant systems or capabilities including existing ones. With a view to ensuring compliance with international law – the very topic of this paper – such an examination may contribute to our understanding of the challenges that autonomy in a weapon system may raise and permit the identification of practical measures, best practices and standards.

29. Against this background, Switzerland suggested in 2016 to describe autonomous weapon systems as *"weapons systems that are capable of carrying out tasks governed by IHL in partial or full replacement of a human in the use of force, notably in the targeting cycle"*. Such a working definition is inclusive, accounts for a wide array of system

[15] Article 4 et seq. of the Articles on State Responsibility for Internationally Wrongful Acts (2001). See also Article 91 of Additional Protocol I to the Geneva Conventions.

configurations, and allows for a debate that is differentiated, compliance-based, and without prejudice to the question of appropriate regulatory response. Indeed, the proposed working definition is not conceived in any way to single out only those systems which could be seen as legally objectionable or otherwise requiring regulation. At one end of the spectrum of systems falling within that working definition, States may find some subcategories to be entirely unproblematic, while at the other end of the spectrum, States may find other subcategories unacceptable. As discussions advance, this working definition needs to evolve to become more specific and purposeful.

IV. A compliance-based approach

30. The present working paper has sought to map the most relevant IHL obligations applicable to the development and employment of autonomous weapon systems on the basis of a broad, inclusive working definition that allows for a discussion of different types of increasingly autonomous weapon systems. On one end of the spectrum of the proposed working definition, some types of autonomous weapon systems would be already unlawful under existing IHL, while on the other end of the spectrum, some types of autonomous weapon systems can be readily qualified as unproblematic.

31. Given the consensus that existing international law applies to all weapon systems, including autonomous weapon systems, and that it has to be respected in all circumstances, an IHL "compliance-based" approach presents itself as a way forward for the CCW discussions. The approach would have three main parts:

32. Firstly, in order to advance our understanding of autonomy and its relation to IHL, States could, based on the proposed working definition, *assess existing autonomous weapon systems and existing systems with* limited autonomy in the targeting cycle. The specific parameters which make a particular system IHL compliant could be identified and examined. These could then be extrapolated to future systems with higher levels of autonomy, to gain an understanding of what features contribute to conformity with or – conversely – objectionability under IHL.

33. Secondly, *reaffirm and spell out applicable international law, in particular IHL*. While many provisions of IHL are well known in principle, there would seem to be merit in collating and clarifying, for ease of reference, the relevant existing provisions as they apply to autonomous weapon systems . Such an exercise would involve the three angles treated in this working paper, i.e. the substantive IHL provisions, legal reviews, and accountability. This working paper has also put forward the notion that – given the current state of robotics and artificial intelligence – the relevant question is not whether a certain level of human control is called for, but what kind and level of human involvement in each of the different phases ranging from conceptualization, development and testing, to operational programming, employment and target engagement. At the heart of the issue is the question: what is the right quality of the human-machine interaction to ensure and facilitate compliance with IHL?

34. Thirdly, still with a view to securing and facilitating compliance, *identify best practices, technical standards and policy measures* that, rather than being strictly part of applicable international obligations, complement, promote and reinforce them. For example, with regard to legal reviews, it could be of interest to identify standard methods and protocols for testing autonomous weapon systems that take into account their unique characteristics. In the same vein, should gaps in the chain of accountability become apparent, States would potentially want to discuss complementary or additional means of ensuring that individual accountability is maintained.

35. The above is not meant to exhaust the debate and is without prejudice to *contemplating further regulatory responses* as may be deemed required. Indeed, discomfort has been expressed by several States and civil society organizations that the delegation of "life and death" decisions to machines is unacceptable and would run against the principles of humanity and the dictates of public conscience. This highlights the importance of human dignity as one of the core values of our international community and as a key element underpinning many international instruments, notably the sources of IHL. While the

question whether there should be limits to handing over certain functions to autonomous weapon systems is a legitimate and important discussion that has its place, it is a question that is very different from the one of compliance with existing international law.

36. In this sense, the "compliance-based" approach submitted here is meant to be part and parcel of a broader "building block" approach that gives space for legal, military, ethical and other legitimate considerations to inform discussions on possible regulatory responses. There appears to be a continued interest in addressing all relevant aspects relating to autonomous weapon systems, notably technological definitions, military utility as well as legal and ethical aspects, under involvement of a variety of stakeholders.

37. With regard to CCW process itself, the work of the Governmental Group of Experts (GGE) provides the opportun y to consolidate the broadly shared understanding that the application of, and compliance with, International Law and in particular International Humanitarian Law. There would be merit to continue studying, based on a broad working definition, all aspects related to the increasing autonomy in weapon systems and, in accordance with the approach presented in this paper, include work along the three main areas of activity mentioned above.